R Worksheet #2

R data management, functions, advanced plots and graphs

Dr. Scott

CSI 500

Spring 2019

# Problem #1: R Data frame management

In the lectures, we discussed how to manage using data frames. Use your R skills to accomplish the following.

## Create R Frames

Given the following data, create R data frames, and then provide R code to create a combined frame that merges these frames by joining on the common ID.

|  |  |  |
| --- | --- | --- |
| id | major | GPA |
| 123 | Data Science | 3.8 |
| 245 | Pre Med | 3.2 |
| 387 | Public Health | 3.4 |

|  |  |  |
| --- | --- | --- |
| id | dorm | in\_state |
| 123 | N | Y |
| 245 | Y | Y |
| 387 | Y | N |

## Frame selection

Write R code to select only the **major** and **in\_state** fields from the combined data frame.

# Problem #2: Functions

Earlier in the course, you developed Python code to compute the value of the Fibonacci sequence. Use your R skills to create a function in R that computes the first 20 values of the Fibonacci sequence. You may wish to use the printf() method discussed in the videos to nicely format your output.

# Problem #3: Scientific Analysis and Visualization

In the lectures, we discussed various ways of rendering multiple data sets for visualization. Use your R skills to analyze the following data and prepare a scientific graph.

# Import the *Modern Applied Statistics with S+* (MASS) library into your R environment

library( MASS )

# From the MASS library, make a copy of the crabs data set

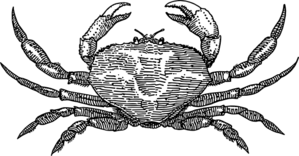
crab.data = MASS::crabs

Write R code to do the following.

## 3.1 Add a new computed column

Add a new column named *wl.ratio* to the crab.data data frame that expresses the ratio of crab carapace length to crab body length. The carapace width is expressed in the CW column, and the body length is expressed in the CL column.

carapace width



carapace length

Image: courtesy [www.clker.com](http://www.clker.com)

## 3.2 Create subset data frames

Create four new data frames, using the R data frame subsetting method of your choice, that divide the data up as follows. Provide R statements to extract the data into the new frames.

species = B, sex = M

species = B, sex = F

species = O, sex = M

species = O, sex = F

## 3.3 Create a figure

Create a single boxplot that displays the *wl.ratio* for each of the data frames you just created. Include a main title, and include appropriate labels for each plot element using the names() command. Your result should look something like this.

